

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 17:36:58 ; Search time 5616 Seconds

(without alignments)  
11345.132 Million cell updates/sec

Title: US-09-725-010-1

Sequence: 1 ggaattcttcttctcattc.....ctttagtttcgcaataatg 1470

Scoring table:

IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: 1: gb ba: 2: gb htg: 3: gb in: 4: gb om: 5: gb ov: 6: gb pac: 7: gb ph: 8: gb pl: 9: gb pr: 10: gb ro: 11: gb sts: 12: gb sy: 13: gb un: 14: gb vi: 15: em ba: 16: em fun: 17: em hum: 18: em in: 19: em mu: 20: em om: 21: em or: 22: em ov: 23: em pac: 24: em ph: 25: em pl: 26: em ro: 27: em sts: 28: em un: 29: em vi: 30: em htg hum: 31: em htg inv: 32: em htg other: 33: em htg mus: 34: em htg pin: 35: em htg rod: 36: em htg vrc: 37: em htg vrc: 38: em sy: 39: em htg hum: 40: em htg mus: 41: em htgo other:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1470	100.0	1470	6	AX154641	AX154641 Sequence
2	505	34.4	2682	8	CA064206	U64206 Candida alb
3	153	10.4	2188	8	AF001978	AF001978 Candida a
4	136.2	9.3	42565	8	CAC35A5	AL033336 C.albican
5	91.2	6.2	86826	3	PFMAL3P5	AL034556 Plasmodiu
6	86.8	5.9	104992	2	AC005504	AC005504 Plasmodiu
7	86.8	5.9	169546	2	AC004157	AC004157 Plasmodiu
8	86.8	5.9	250421	3	AE014849	AE014849 Plasmodiu
9	77.6	5.3	105682	3	AC116977_3	Continuation (4 of
10	75.2	5.1	302156	3	AC116977	AC116977 Dictyoste
11	74.6	5.1	343050	3	PFA929353	AF929353 Plasmodiu
12	74	5.0	76568	3	MBREV	AF538053 Monosiga
13	73.8	5.0	250663	3	AE014826	AE014826 Plasmodiu
14	73.4	5.0	14867	3	AE001398	AE001398 Plasmodiu
15	73.4	5.0	164399	3	PFMAL3P6	Z98551 Plasmodiu
16	72.8	5.0	80803	9	AC133522	AC133522 Homo sapi
17	72.8	5.0	113880	3	PFMAL3P4	AL008970 Plasmodiu
18	72.6	4.9	110000	2	PFMAL7P1_07	Continuation (8 of
19	72	4.9	8056	6	AX599046	AX599046 Sequence
20	72	4.9	349751	3	PFMAL4P3	AL035476 Plasmodiu
21	71.8	4.9	145892	9	AC134919	AC134919 Homo sapi
22	71.8	4.9	343050	3	PFA929353	AF929353 Plasmodiu
23	71.4	4.9	25568	3	AC116955	AC116955 Dictyoste
24	71.4	4.9	39984	3	AC116101	AC116101 Dictyoste
25	70.8	4.8	2009	6	AX457067	AX457067 Sequence
26	70.8	4.8	271546	3	AE014843	AE014843 Plasmodiu
27	70.4	4.8	192929	2	AC005505	AC005505 Plasmodiu
28	70.4	4.8	250713	3	AE014850	AE014850 Plasmodiu
29	70.2	4.8	281723	3	PFA929359	AF929359 Plasmodiu
30	69.4	4.7	182870	3	AC116960	AC116960 Dictyoste
31	69.2	4.7	253305	3	PFMAL3P7	AL034559 Plasmodiu
32	68.6	4.7	166719	9	AC136957	AC136957 Homo sapi
33	68.6	4.7	198434	9	AC109994	AC109994 Homo sapi
34	68.2	4.6	192187	3	AC116920	AC116920 Dictyoste
35	68	4.6	9810	6	AX45328	AX45328 Sequence
36	67.8	4.6	250029	3	AE014830	AE014830 Plasmodiu
37	67.6	4.6	170966	5	BX000991	BX000991 Zebrafish
38	67.4	4.6	66993	2	AC138074	AC138074 Homo sapi
39	67.2	4.6	1708	8	AF500786	AF500786 Plasmocys
40	67.2	4.6	7347	1	AF211124	AF211124 Carsonell
41	67.2	4.6	67970	3	PFMAL3P1	AL031746 Plasmodiu
42	67.2	4.6	110000	2	PFMAL3P1_04	Continuation (5 of
43	67	4.6	8056	6	AX599046	AX599046 Sequence
44	67	4.6	110000	2	AL954295_2	Continuation (3 of
45	67	4.6	122838	2	BX248494	BX248494 Danio rer

## ALIGNMENTS

RESULT 1  
AX154641  
LOCUS AX154641 1470 bp DNA linear PAT 22-JUN-2001  
DEFINITION Sequence 1 from Patent WO0138550.  
ACCESSION AX154641  
VERSION AX154641.1 GI:14536200  
KEYWORDS  
SOURCE  
ORGANISM  
Candida albicans  
Bukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitosporic Saccharomycetales; Candida.  
REFERENCE  
AUTHORS Sundstrom, P.  
TITLE Methods for altering the expression of hyphal-specific genes

FEATURES  
Source Sundstrom, Paula (US)  
Location/Qualifiers  
1. 1470  
/organism="Candida albicans"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:5476"

## ORIGIN

Query Match 100.0%; Score 1470; DS 6; Length 1470;  
Best Local Similarity 100.0%; Pred. No. 3.6e-226;  
Matches 1470; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGATCTTTCTTTTCATTTCCCTTAAACCGATCAGAAAGAAAGTGAATTAAGCTAT 60  
DB 1 GGATCTTTCTTTTCATTTCCCTTAAACCGATCAGAAAGAAAGTGAATTAAGCTAT 60  
QY 61 GATTAATGTTGATTTGTTGATTAATCAATCACTAAGACGTTGACAGTTAAAGTACG 120  
DB 61 GATTAATGTTGATTTGTTGATTAATCAATCACTAAGACGTTGACAGTTAAAGTACG 120  
QY 121 TTGTTGTGTCCTGCTCGTCTAATTTCTGTTGACAGATTAATTAACAAGAAATACG 180  
DB 121 TTGTTGTGTCCTGCTCGTCTAATTTCTGTTGACAGATTAATTAACAAGAAATACG 180  
QY 181 GAAACCCCTCCAAAATAAATTTTGAACCTTACAGCAGCAATTAATGCGATTAACCTTCG 240  
DB 181 GAAACCCCTCCAAAATAAATTTTGAACCTTACAGCAGCAATTAATGCGATTAACCTTCG 240  
QY 241 CATATATAAATCTTTGAAACATACGATATGTTATTTCTTTCAATCTGAAATATTTT 300  
DB 241 CATATATAAATCTTTGAAACATACGATATGTTATTTCTTTCAATCTGAAATATTTT 300  
QY 301 GCTTTTCTTTTAACTTATGAACAATTAAGAAAGAAAGTGAAGTGAAGTGC 360  
DB 301 GCTTTTCTTTTAACTTATGAACAATTAAGAAAGAAAGTGAAGTGAAGTGC 360  
QY 361 TAACCATGAAATATATAGGCTTAAGGTTTTCCTGATCGTTTAACTAAAAGGAAATAA 420  
DB 361 TAACCATGAAATATATAGGCTTAAGGTTTTCCTGATCGTTTAACTAAAAGGAAATAA 420  
QY 421 CAAAAGTTATAGGATACCTGCTGAAGTGTCAACAAATATATTTTGAACGTTAGCT 480  
DB 421 CAAAAGTTATAGGATACCTGCTGAAGTGTCAACAAATATATTTTGAACGTTAGCT 480  
QY 481 CCAATGAAATATATCAAACTTAATCTTAAGAAATTCCTATATATATATAGGAAATCC 540  
DB 481 CCAATGAAATATATCAAACTTAATCTTAAGAAATTCCTATATATATATAGGAAATCC 540  
QY 541 CTCTCAGCTGACTGAATATATCCATCTGAATTAATCACTCACTATATTCATATAA 600  
DB 541 CTCTCAGCTGACTGAATATATCCATCTGAATTAATCACTCACTATATTCATATAA 600  
QY 601 TAGATTAAGTATGTTCTCTCTCACTAGTACATTAATGCAATGAGTACGTTATG 660  
DB 601 TAGATTAAGTATGTTCTCTCTCACTAGTACATTAATGCAATGAGTACGTTATG 660  
QY 661 TTCAATTAATTAAGCACTGTTGACACCTTAATTCGAACATTAATCTGATCAATTTTCTT 720  
DB 661 TTCAATTAATTAAGCACTGTTGACACCTTAATTCGAACATTAATCTGATCAATTTTCTT 720  
QY 721 GTCCTCTCTGTTGTTTCTCTCACTAGTACATTAATGCAATGAGTACGTTATG 780  
DB 721 GTCCTCTCTGTTGTTTCTCTCACTAGTACATTAATGCAATGAGTACGTTATG 780  
QY 781 AACACATTAACCTTTGAGTATGATTAATTAATCACTTAATGCTTTTGAAGTAAAGAT 840  
DB 781 AACACATTAACCTTTGAGTATGATTAATTAATCACTTAATGCTTTTGAAGTAAAGAT 840  
QY 841 CAAATTTTCTTAACTGACTTAAGCACTTATGCAATGAGTATGATTTGATCTAC 900  
DB 841 CAAATTTTCTTAACTGACTTAAGCACTTATGCAATGAGTATGATTTGATCTAC 900  
QY 901 TACTATAAGCTCAACAAATTAATCTTTCAAAAATGTTAATTAACAAGTCACTTAAT 960

DB 901 TACTATAAGCTCAACAAATTAATCTTTCAAAAATGTTAATTAACAAGTCACTTAAT 960  
QY 961 TCTTGGATCCCAAAACAAGAAATTCGAAATTCGAGATTAATGTCAGTCACAATTC 1020  
DB 961 TCTTGGATCCCAAAACAAGAAATTCGAAATTCGAGATTAATGTCAGTCACAATTC 1020  
QY 1021 ATTGTAAGGAGGAGTTTGTGTAGGCTCAATTCGCTTATATAGTACCTTAAGTAA 1080  
DB 1021 ATTGTAAGGAGGAGTTTGTGTAGGCTCAATTCGCTTATATAGTACCTTAAGTAA 1080  
QY 1081 TCTAATAACAACAACCTTTCTAATAACCTTAATTAATTAATTAATTAATTAATTAAT 1140  
DB 1081 TCTAATAACAACAACCTTTCTAATAACCTTAATTAATTAATTAATTAATTAATTAAT 1140  
QY 1141 GATTAAGTATGTTAGCCGCTGTTTCTTTTGTGCTTATTTTATGATCACTTTGTT 1200  
DB 1141 GATTAAGTATGTTAGCCGCTGTTTCTTTTGTGCTTATTTTATGATCACTTTGTT 1200  
QY 1201 CACTTTTGTGCACTTATTAATACCTTTTGAACACTCTTTGATACCTGATACCG 1260  
DB 1201 CACTTTTGTGCACTTATTAATACCTTTTGAACACTCTTTGATACCTGATACCG 1260  
QY 1261 CCTTTTCTTAACTAGCACTTTGTAAGTCCCTTTCTTTTCCACTAATTTTATCATCT 1320  
DB 1261 CCTTTTCTTAACTAGCACTTTGTAAGTCCCTTTCTTTTCCACTAATTTTATCATCT 1320  
QY 1321 TGAATATGTAATCAATAGTTTCAAAAATTAATTAATTAATTAATTAATTAATTAAT 1380  
DB 1321 TGAATATGTAATCAATAGTTTCAAAAATTAATTAATTAATTAATTAATTAATTAAT 1380  
QY 1381 ATTTCATTTCCATTCATCACTTTTCTCAACATTAATCAACACAGAAATCTCTTA 1440  
DB 1381 ATTTCATTTCCATTCATCACTTTTCTCAACATTAATCAACACAGAAATCTCTTA 1440  
QY 1441 TAGTCACTGCTTTTGTGTTGCTGATATG 1470  
DB 1441 TAGTCACTGCTTTTGTGTTGCTGATATG 1470

RESULT 2  
CAU64206 2682 bp DNA linear PLN 17-DEC-1998  
LOCUS Candida albicans hyphal wall protein 1 (HWP1) gene, complete cds.  
DEFINITION U64206  
VERSION U64206.1 GI:4028879  
KEYWORDS  
SOURCE  
ORGANISM  
Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Microsporica; Saccharomycetales; Candida.  
REFERENCE  
1 (bases 1 to 2682)  
Staab,J.F., Ferrer,C.A. and Sundstrom,P.  
Developmental expression of a tandemly repeated, proline-and  
glutamine-rich amino acid motif on hyphal surfaces on Candida  
albicans  
JOURNAL  
J Biol. Chem. 271 (11), 6298-6305 (1996)  
MEDLINE  
66198091  
PUBMED  
8626424  
REFERENCE  
2 (bases 1 to 2682)  
Staab,J.F. and Sundstrom,P.  
Genetic organization and sequence analysis of the hypha-specific  
cell wall protein gene HWP1 of Candida albicans  
JOURNAL  
Yeast 14 (7), 681-686 (1998)  
MEDLINE  
98301186  
PUBMED  
9639315  
REFERENCE  
3 (bases 1 to 2682)  
Staab,J.F. and Sundstrom,P.  
Direct Submision  
Submitted (18-JUL-1996) Medical Microbiology & Immunology, Ohio  
State University, 333 West 10th Avenue, Columbus, OH 43210, USA  
4 (bases 1 to 2682)  
Staab,J.F.  
REFERENCE  
AUTHORS



```

RESULT 4
CAC35A5/c 42565 bp DNA linear PLN 05-NOV-1998
DEFINITION C.albicans cosmid Ca35A5.
ACCESSION AL033396
VERSION AL033396.1 GI:3850143
KEYWORDS ARS; Canik1; CDP-alcohol phosphatidyltransferase; chitin synthase;
chsi; cytochrome P450; DNA polymerase delta; histidine kinase;
isocitrate dehydrogenase pseudogene; LTR; pol3; Rcol; regulator of
chromosome condensation; rehydrin; Rps1-like region.
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
1 (bases 1 to 42565)
Tait, B., Simon, M.C., King, S., Brown, A.J., Gow, N.A. and Shaw, D.J.
A Candida albicans genome project: cosmid contigs, physical
mapping, and gene isolation
Fungal Genet. Biol. 21 (3), 308-314 (1997)
JOURNAL 97435544
MEDLINE 9290243
PUBMED Article No. FG970983
REFERENCE 2 (bases 1 to 42565)
AUTHORS Oliver, K. and Harris, D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 42565)
AUTHORS Barrell, B.G. and Randal, M.A.
JOURNAL Direct Submission
Submitted (05-NOV-1998) On behalf of the pilot sequencing project
on the Candida albicans strain 161 genome. Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA E-mail:
barrell@sanger.ac.uk cosmids supplied by Prof. Duncan Shaw [3]
Department of Molecular and Cell Biology, The Institute of Medical
Science, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD,
UK
COMMENT Notes:
Funding: sequencing funded by Beowulf Genomics Ltd. CDS are
numbered using the following system eg CAC20C1.01C. CA (C.
albicans), 20C1 (cosmid name), 01 (complementary strand).
The more significant matches with motifs in the PROSITE/PRM
database are also included but some of these may be fortuitous.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid Ca35A5 is
likely to map to region R, chromosome 7.
location/Qualifiers
1..42565
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="1161"
/db_xref="taxon:5476"
/map="chromosome 7"
/clone="cosmid Ca35A5"
/complement(1..785)
/gene="Ca35A5.01c"
/complement(1..785)
/gene="Ca35A5.01c"
/more="Ca35A5.01c, partial orf, len: > 261 aa, most
similar to EPT1_YEAS ethanolaminephosphotransferase (EC
2.7.8.1) (391 aa), fasta scores opt: 984, E(1): 0, (56.3%
identity in 263 aa overlap), contains P500379 CDP-alcohol
phosphatidyltransferases signature"
/codon_start=1
/transl_table=12
/product="putative alcohol phosphatidyl transferase"
/protein_id="CAA21944.1"
/db_xref="GI:3850144"
/db_xref="GOA:O94010"
/db_xref="SPTREMBL:O94010"
/translation="MGILFIPNKLQNLKLYKYSSEDSHSISKYLTKKWNPFVQIFPL

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 17:33:53 ; Search time 593 seconds

(without alignments)  
10530.946 Million cell updates/sec

Title: US-09-725-010-1

Perfect score: 1470  
Sequence: 1 ggaatctctcttcattc.....ctttgagttcgtcaatcg 1470

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1980s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1459	99.3	1471	5 AAF90471	Aaf90471 Candida a
2	159.4	10.8	2093	3 AAF57996	Aaf57996 2093 bp C
3	72	4.9	8036	7 ABZ10246	Abz10246 Haematopo
4	68	4.6	9810	6 ABZ10246	Abz10246 Human imm
5	67	4.6	8056	7 ABZ10246	Abz10246 Haematopo
6	66.4	4.5	18624	6 ABZ13702	Abz13702 Human imm
7	66	4.5	17538	6 ABZ13702	Abz13702 Human imm
8	65.6	4.5	7584	6 AAS95251	Aas95251 Long term
9	64.8	4.4	5884	6 ABZ13165	Abz13165 Human imm
10	64.4	4.4	34688	6 ABZ10246	Abz10246 Human imm
11	64	4.4	5254	6 ABZ13165	Abz13165 Human imm
12	63.4	4.3	115218	7 ACA64845	ACA64845 Signal tr
13	62.8	4.3	15121	6 ABN80239	Abn80239 Human HNR
14	62	4.2	8056	7 ABZ10100	Abz10100 Haematopo
15	61.2	4.2	6775	6 ABO67159	Abg67159 Human ang
16	61.2	4.2	12592	6 AAS61102	Aas61102 Human gen
17	61	4.1	6620	4 AAS4487	Aas4487 Chemically
18	61	4.1	6620	6 ABK28415	ABK28415 DNA trans
19	61	4.1	6620	6 ABN80278	Abn80278 Human che
20	61	4.1	8056	7 ABZ10100	Abz10100 Haematopo
21	60.2	4.1	12426	4 AAS46482	Aas46482 Tumour su
22	60	4.1	1501	7 ABZ10188	Abz10188 Haematopo
23	60	4.1	1501	9 ADE84162	Ade84162 Human lym

24	59.8	4.1	8310	2 AAZ29911	Aaz29911 cDNA enco
25	59.6	4.1	6494	6 ABZ13392	Abz13392 Human imm
26	59.6	4.1	6494	6 AAD28390	Aad28390 Human che
27	59.2	4.0	3684	7 ACB28280	AcB28280 Colon can
28	59.2	4.0	3684	7 ABZ10128	Abz10128 Haematopo
29	59.2	4.0	6664	9 ADB54322	AdB54322 Pretreate
30	59.2	4.0	11334	6 ABZ13233	Abz13233 Human imm
31	59.2	4.0	18154	6 ABZ13255	Abz13255 Human imm
32	58.8	4.0	10048	6 ABZ170314	Abz170314 Chemical
33	58.8	4.0	10048	6 AAS61252	Aas61252 Human gen
34	58.6	4.0	15373	6 ABZ132467	Abz132467 Human imm
35	58.4	4.0	17131	6 ABZ13053	Abz13053 Human imm
36	58.2	4.0	11049	6 ABZ132669	Abz132669 Human imm
37	58.2	4.0	11049	6 ABZ192219	Abz192219 Chemical
38	58.2	4.0	11049	6 ABZ19322	Abz19322 Human pol
39	58	3.9	5474	6 ABZ133041	Abz133041 Human imm
40	58	3.9	5474	6 ABO66992	Abg66992 Human ang
41	58	3.9	7631	6 ABZ132860	Abz132860 Human imm
42	58	3.9	13574	6 ABZ13317	Abz13317 Human imm
43	57.8	3.9	5085	7 ACB62814	AcB62814 Colon can
44	57.8	3.9	8085	9 ADB54280	AdB54280 Pretreate
45	57.4	3.9	883	4 AAL15210	Aal15210 Human bre

## ALIGNMENTS

RESULT 1	AAf90471	standard; DNA; 1471 bp.
ID	AAf90471	
XX	AAf90471:	
AC	AAf90471:	
DT	06-AUG-2001 (first entry)	
XX		
DE	Candida albicans HWP1 gene 5' upstream region.	
XX		
KM	HypH1 wall protein; HWP1; Promoter; hypH1-specific gene; infection;	
KX	candidiasis; therapy; antifungal; fungicide; ds.	
XX		
OS	Candida albicans.	
XX		
FH	Key	Location/Qualifiers
FT	Promoter	1..1468
FT		/*tag= u
FT	protein_bind	complement(59..64)
FT		/*tag= a
FT	protein_bind	/bound_moiety= "NIT2"
FT		94..105
FT		/*tag= b
FT	protein_bind	/bound_moiety= "PHO4"
FT		complement(228..233)
FT		/*tag= c
FT	protein_bind	/bound_moiety= "NIT2"
FT		complement(265..270)
FT		/*tag= d
FT	protein_bind	/bound_moiety= "NIT2"
FT		complement(432..438)
FT		/*tag= e
FT	protein_bind	/bound_moiety= "NIT2"
FT		468..479
FT		/*tag= f
FT	protein_bind	/bound_moiety= "PHO4"
FT		470..478
FT	misc_feature	/*tag= g
FT		/function= "S.cerevisiae cAMP response element"
FT		560..565
FT	protein_bind	/*tag= h
FT		/bound_moiety= "NIT2"
FT		573..578
FT	protein_bind	/*tag= i
FT		/bound_moiety= "NIT2"
FT		705..710
FT	protein_bind	

```

protein_bind
/*tag= j
/bound_moiety= "NIT2"
complement(800. .805)
/*tag= k
/bound_moiety= "NIT2"
807. .812
/*tag= l
/bound_moiety= "NIT2"
921. .926
/*tag= m
/bound_moiety= "NIT2"
1139. .1144
/*tag= n
/bound_moiety= "NIT2"
1146
/*tag= o
/note= "this base is not present in the HWP1 gene 5'
flanking region sequence given in Fig 4 of the
specification"
1247. .1252
/*tag= p
/bound_moiety= "NIT2"
1256. .1261
/*tag= q
/bound_moiety= "NIT2"
1313. .1318
/*tag= r
/bound_moiety= "NIT2"
1355. .1361
/*tag= t
1422. .1427
/*tag= s
/bound_moiety= "NIT2"
WO200138550-A2.
31-MAY-2001.
29-NOV-2000; 2000WO-US032464.
29-NOV-1999; 99US-0167672P.
(SUND/) SUNDSTROM P.
Sundstrom P;
WPI, 2001-367698/38.
interfering with expression of hyphal-specific genes in fungus for
inhibiting fungal cell growth involves interfering with transcription of
hyphal-specific genes mediated by cis acting sequences.
claim 21; Page 78-89; 95pp; English.

```

CC diabetic ketoacidosis, and patients in which the normal microbial flora  
CC has been disrupted because of disease, trauma or chemical, radiation or  
CC other immunosuppressive prophylaxis. Also provided is a method for  
CC characterising genes under control of a DNA binding protein  
XX  
XX Sequence 1471 BP, 509 A; 260 C; 186 G; 516 T; 0 U; 0 Other;

Query Match	99.3%	Score 1459;	DB 5;	Length 1471;
Best Local Similarity	99.9%	Pred. No. 3.1e-258;		
Matches 1470; Conservative	0;	Mismatches 0;	Indels 1;	Gaps 1

Qy	60	GGATCTTTCTTTCACTTCCCTTAAACCGATCAAGAAAGAAAGTGAATTAAGCTAT	60
Db	1	GGATCTTTCTTTCACTTCCCTTAAACCGATCAAGAAAGAAAGTGAATTAAGCTAT	60
Qy	61	GATTAATGTTGATTTTGTGTATTCATCAACTAGACGCTTTCACGTTAAAAAGTAGC	120
Db	61	GATTAATGTTGATTTTGTGTATTCATCAACTAGACGCTTTCACGTTAAAAAGTAGC	120
Qy	121	TTGTGTGTGCTCGCTCGCTGCTAATTTCTGTACACAGAGTAAATATACAAAGAAATACAG	180
Db	121	TTGTGTGTGCTCGCTCGCTGCTAATTTCTGTACACAGAGTAAATATACAAAGAAATACAG	180
Qy	181	GAATCCCTCAAAAAAAATTTTGGACCTTACACGACATTAATTGCGATTAACCTTGC	240
Db	181	GAATCCCTCAAAAAAAATTTTGGACCTTACACGACATTAATTGCGATTAACCTTGC	240
Qy	241	CATAATAAAAACCTTTGATTAACATACACGATATGTATCTTTTCACTAGGAAATATTTT	300
Db	241	CATAATAAAAACCTTTGATTAACATACACGATATGTATCTTTTCACTAGGAAATATTTT	300
Qy	301	GCTTTTATTAAACATTAATGACATTTGAGAAAAAAGAAAAATGAAGTAAGTAGTTC	360
Db	301	GCTTTTATTAAACATTAATGACATTTGAGAAAAAAGAAAAATGAAGTAAGTAGTTC	360
Qy	361	TAAACCTTAAATTAATTAAGCTTAAGCTTTTTCGATGCGCTTAACTTAAAAAGAAATTA	420
Db	361	TAAACCTTAAATTAATTAAGCTTAAGCTTTTTCGATGCGCTTAACTTAAAAAGAAATTA	420
Qy	421	CAAAAGTTATTAAGCATTAACCTGCGTAAAGTGTCAACAAATATATTTTGCAGCTTAC	480
Db	421	CAAAAGTTATTAAGCATTAACCTGCGTAAAGTGTCAACAAATATATTTTGCAGCTTAC	480
Qy	481	CTTATGAAAAATTAACAACTAAATCTTAAAGAAATTTCTCTATATATATTAAGAAATCC	540
Db	481	CTTATGAAAAATTAACAACTAAATCTTAAAGAAATTTCTCTATATATATTAAGAAATCC	540
Qy	541	CTCTCAGATGAACGTAAATTCATCGAATTAACGTCCACATTAATTCATCATATAAA	600
Db	541	CTCTCAGATGAACGTAAATTCATCGAATTAACGTCCACATTAATTCATCATATAAA	600
Qy	601	TAGATTAGTATGTTCTCTTCAAGTACATTACTACATTAATGCAATGCTAGCTATTC	660
Db	601	TAGATTAGTATGTTCTCTTCAAGTACATTACTACATTAATGCAATGCTAGCTATTC	660
Qy	661	TTGATTAATGACATGTTGACACACCTTAATTCGAACTTAACTGATACCATATTTTCTT	720
Db	661	TTGATTAATGACATGTTGACACACCTTAATTCGAACTTAACTGATACCATATTTTCTT	720
Qy	721	GTCCTTCTTTGTTTTTTCTAACAAAAATGTCAGAAATTTTTTAAAAAATTTGAAA	780
Db	721	GTCCTTCTTTGTTTTTTCTAACAAAAATGTCAGAAATTTTTTAAAAAATTTGAAA	780
Qy	781	AACACATAACCTTGAATGATATATATCAACTATTAAGCTGTGTAAGTAAGTAAGAT	840
Db	781	AACACATAACCTTGAATGATATATATCAACTATTAAGCTGTGTAAGTAAGTAAGAT	840
Qy	841	CAAAATTTTTTTCTAACTGACCTAATGCACTTTACATCAACTGATGTAATTTGCATTCAC	900
Db	841	CAAAATTTTTTTCTAACTGACCTAATGCACTTTACATCAACTGATGTAATTTGCATTCAC	900
Qy	901	TACTATTAAGCTAAACAAATTAATCTTTCAAAAAAGTATATTAATTAACAGCATATTAAT	960

Db 901 TACTATAGCTAAACAAATTAATCTTTCAAAAAGTATATTAACAGTCACTATAT 960  
 QY 961 TCTTTGATCCAAAAACAAGAAATTCGAAATTCGAGCAATTAATGTCACATATTC 1020  
 Db 961 TCTTTGATCCAAAAACAAGAAATTCGAAATTCGAGCAATTAATGTCACATATTC 1020  
 QY 1021 ATTTGAAAAAGGAGGTTTGGGCTCATATGCTTATATGTAAGTAAAGTAA 1080  
 Db 1021 ATTTGAAAAAGGAGGTTTGGGCTCATATGCTTATATGTAAGTAAAGTAA 1080  
 QY 1081 TCTAAACCAACACAACTTTCTAAACCTATATATATACCTTAATGCTCAACCGG 1140  
 Db 1081 TCTAAACCAACACAACTTTCTAAACCTATATATATACCTTAATGCTCAACCGG 1140  
 QY 1141 GATTA-GTTAGTAAAGGAGGCTTTTGGCTTATATGTAAGTAAAGTAA 1199  
 Db 1141 GATTAAGTAAAGTAAAGGAGGCTTTTGGCTTATATGTAAGTAAAGTAA 1200  
 QY 1200 TCACTTTTGGTGGCACTTAAATACCGTTTGGCACTTCTGTAATCACTGATCC 1259  
 Db 1201 TCACTTTTGGTGGCACTTAAATACCGTTTGGCACTTCTGTAATCACTGATCC 1260  
 QY 1260 GCGTTTAAACAGACAACTCTGTAAGTCCCTTTTCCCACTATTTATCATTC 1319  
 Db 1261 GCGTTTAAACAGACAACTCTGTAAGTCCCTTTTCCCACTATTTATCATTC 1320  
 QY 1320 TGAATATATGTAATGAGATTTTTCAAAACTATATATAGGTCAAAAATACCGG 1379  
 Db 1321 TGAATATATGTAATGAGATTTTTCAAAACTATATATAGGTCAAAAATACCGG 1380  
 QY 1380 TATTTCAATTTCCATTCACCTGTTTCTCAACATATCAACACAGAAATCTCT 1439  
 Db 1381 TATTTCAATTTCCATTCACCTGTTTCTCAACATATCAACACAGAAATCTCT 1440  
 QY 1440 ATAGTCACCTGCTTAACTTGTGCTCAATNG 1470  
 Db 1441 ATAGTCACCTGCTTAACTTGTGCTCAATNG 1471

RESULT 2  
 ID AA57996/C  
 AC AA57996 standard; DNA; 2093 BP.  
 AC AA57996;  
 DT 10-OCT-2000 (first entry)  
 DE 2093 bp Candida albicans retrotransposon 15 sequence.  
 KM Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KM gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KM reverse transcriptase; RNaseH; pseudonoc; readthrough translation;  
 KM stop codon suppression; gene delivery; gene therapy vector;  
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.  
 OS Candida albicans.  
 XX WO200026397-A1.  
 XX 11-MAY-2000.  
 XX PD 01-NOV-1999; 99WO-NZ000179.  
 XX PF 30-OCT-1998; 98CA-02249046.  
 XX PR 30-OCT-1998; 98DS-0106342P.  
 XX (JANC) JANSSEN PHARM NV.  
 XX PA Luyten WHM, De Backer MD, Nelissen BJM, Poulter RTM,  
 XX PI KPI; 2000-365640/31.  
 XX DR Novel retrotransposon expression vectors useful for expressing an  
 XX PT

PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 of Candida in a sample.  
 PS Disclosure; Fig 71; 204pp; English.  
 XX The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pCal (AA57992), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
 CC H031042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudonoc within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57939-457966 and  
 CC AA57982-458019 represent novel C. albicans retrotransposon sequences

Sequence 2093 BP; 747 A; 332 C; 302 G; 712 T; 0 U; 0 Other;  
 SQ

Query Match 10.8%; Score 159.4; DB 3; Length 2093;  
 Best Local Similarity 84.9%; Pred. No. 4.4e-20;  
 Matches 191; Conservative 0; Mismatches 31; Indels 3; Gaps 1;

QY 480 TCTATAGAAATATACAACTTAATCTTAAAGAAATTCCTCTATATATATAGAAATC 539  
 Db 1237 TCTATAGAAATATACAACTTAATCTTAAAGAAATTCCTCTATATATATAGAAATC 1178  
 QY 540 CCTCTCAGAGGAGCTGATATATCATCTGATTAATATGCTCAGTAATTC--CATCAAT 596  
 Db 1177 CCTCTCAGAGGAGCTGATATATCATCTGATTAATATGCTCAGTAATTC--CATCAAT 1118  
 QY 597 AAATATGATATGATATATGTTCTCTCAATCAATTAATCAATTAATCAATTAATCAAT 656  
 Db 1117 GGAATATGATATGATATATGTTCTCTCAATCAATTAATCAATTAATCAATTAATCAAT 1058  
 QY 657 ATTTGTAATATGTCAGATGTTGACACCCCAATTCGAACATTAA 701  
 Db 1057 ATTTGTAATATGTCAGATGTTGACACCCCAATTCGAACATTAA 1013

RESULT 3  
 ID AB210246/C  
 AC AB210246 standard; DNA; 8056 BP.  
 AC AB210246;  
 DT 16-JAN-2003 (first entry)  
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.  
 KM Human; haematopoietic cell proliferation disorder; cytostatic;  
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KM cytosine methylation state; gene; ds.  
 OS Homo sapiens.  
 XX WO200277272-A2.  
 XX XX  
 XX PN  
 XX PT



PD 03-OCT-2002.  
 XX 26-MAR-2002; 2002WO-EP003401.  
 XX 26-MAR-2001; 2001US-0278333P.  
 PA (EPIC-) EPIGENOMICS AG.  
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J,  
 P1 Olek A, Pispennyock C, Adorjan P, Grabs G, Lesche R, Leu E,  
 P1 Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C,  
 P1 Schwope I, Ziebarth H;  
 DR MPI; 2003-018942/01.  
 XX  
 PT Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent that  
 PT distinguishes between methylated and non-methylated CpG dinucleotides.  
 PS Claim 28; SEQ ID NO 386; 117pp; English.  
 XX  
 XX The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related DNA  
 CC sequences. The nucleotide sequences from the present invention can also  
 CC be used for detecting a predisposition to, differentiation between  
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
 CC haematopoietic cell proliferative disorders. The present method enables a  
 CC highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients  
 CC  
 XX  
 XX  
 SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;  
 Query Match 4.9%; Score 72; DB 7; Length 8056;  
 Best Local Similarity 42.0%; Pred. No. 0.00051;  
 Matches 489; Conservative 0; Mismatches 670; Indels 5; Gaps 1;  
 Oy 203 TTGAACCTTACGACGACATTAATTCGCGATTAATTCGCATATATTAACCTCTTGAAC 262  
 Db 1640 TTTTACAAATTCATTTATTTCAATTTTTTAAATATTAATTAATTAATTAATTAATTAAT 1581  
 Oy 263 ATACGATATGTAATCTTTCTTCAATTAAGTGAATTTTCTCTCTCTTTTAAAGCTA 317  
 Db 1580 ATTAAATTTTAAATATTAATTTTTTAAATTAATTTTAAATTTTAAATTTTAAACAT 1521  
 Oy 318 TGAACATTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 377  
 Db 1520 TTAATTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 1461  
 Oy 378 AGGCTAAGTTTCTCTGATGCTTTAACTTAAAGAAATTAACAAAGCTATTAAGCAT 437  
 Db 1460 AAAAAATTAATTTTAAATTTTTTTTTTAAATTAATTAATTAATTAATTAATTAATTAATC 1401  
 Oy 438 AACCTCGTAAGTGTCACAAAATTAATTTTGAAGCTAGCTATTAAGAAATTAACAA 497  
 Db 1400 AATATTTATTTTAAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTTT 1341  
 Oy 498 AACTAAATCTTAAGAAATTTCTCTATATTAATTAAGAAATCCCTCAAGCTGAAGCA 557  
 Db 1340 TTTATTTAAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTAATTAATTAATTAAT 1281

QY	558	ATTATTCATCGAATTTATGAGTTCAGCTACGATTCATCAATATAAATAGATTTAGTGTATGTT	6117
Db	1280	TTTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	12211
QY	618	CTCTTCAGTACAAATTAATCAATTAATGCAATGCTAGCTATTTGTTCATTAATTAACCGATG	677
Db	1220	AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	11611
QY	678	TGCAACCCCAATTCGACATTAACGTATGATTCGATTTTCTTGCCCTCTTGTTGTTT	737
Db	1160	ATTATATTTTAATTTTCGAATTTTCAAAAAAATTTAAAAAAACAAATTAATTAACATTTTA	11011
QY	738	TCATACAAAAATGTTCCAGAAATTTTAAAAAATTTGAAAAACACATTAACCTTTGA	797
Db	1100	AAACATTAATAAAAAAAACAAATTTAAAAACAAATTTTATTTTACAAATTAATTTTCA	10411
QY	798	GTAAGATTAATTCGAATTAATGACTGTTTGAAAGTAAAGATCAAAATTTTCTTACT	857
Db	1040	AAATTAATAAAAAATTAATAATTTTAAAAATTAATTTTAAAAAAATTAATAATAT	981
QY	858	CGACTTAATGACCTTACATCAACGATGATTTATGTCATCTATCTATTAAGCTTAACA	917
Db	980	ATTTTATTCATTTAAATTAATAAAAAATTTAAATTTTAAAAAAATTTTATTAACAAA	921
QY	918	AATTAATCTTTCAAAAATGCTTAATTAATTAACAGTCAATCTAATCTTTGGATCCAAAAAC	977
Db	920	CACCAAAATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	861
QY	978	AAGAAATTCGAAATTCGACGATTAATGCGACTCACAATTCATTTGTAATAAGGAGAG	1031
Db	860	ATTAATAAAAAACAAATTTTAAAAATTAATAAAAAATTTAAATTAATTAATTAATTAATTTT	801
QY	1038	TTTGTGTGGGCTCATTAATGCTTAATTAATGCTAATGATCTAATAAGATCTAATAACACACAC	1091
Db	800	AAATTAATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	741
QY	1098	CTTCTTAACCACTTAATTAATTAATTAATGCTCAACACGGAGTAAGTTAGTTAGCC	1151
Db	740	TTATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT	681
QY	1158	AGCTGTTTTTTTTGCCCTATTTTAAAGCACTTTGTTGCACTTTTGTCGACT	1211
Db	680	ATTTTATTTTATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	621
QY	1218	TTAATACCGTTTTCGCACTTCCTTGTGTATCACTGTATCCGCTTTTAAACATGACA	1271
Db	620	TTATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	561
QY	1278	ACTCTGTGAAGTCCCTTTCTTTGCCCACTATTTATCATCTTGGAATATGTAATCAGA	1331
Db	560	ATAAATAATTAATTAATAATTTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	501
QY	1338	ATAGTTTTTCAAAAATTAATAATA 1361	
Db	500	ATTTTAAATTTTATTTTAAATA 477	
RESULT 4			
ID	ABL32426/c	ABL32426 standard; DNA; 9810 BP.	
XX	ABL32426;		
XX	26-MAR-2002 (first entry)		
DE	Human immune system associated gene SEQ ID NO: 399.		
XX	Human immune system associated gene SEQ ID NO: 399.		
KM	Human: immune system disease; cytosine methylation; antiasthmatic;		
KM	antibacterioclastic; antianaemic; cytosolic; noctropic;		
KM	neuroprotective; anti-HIV; anticonvulsant; ophthalmological;		
KM	antihemetic; antiarthritic; antidiabetic; antipsoriatic;		
KM	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;		